

## Статистика сайта и ее анализ

---

---

- ⇒ Интернет-статистика: цели и задачи
- ⇒ Какие параметры посещаемости сайта нужно оценивать
- ⇒ Откуда берутся данные для анализа
- ⇒ Обработка данных: собственный анализ данных или счетчик?
- ⇒ Счетчики: возможности, достоинства и недостатки
- ⇒ Обзор возможностей анализаторов журнальных файлов
- ⇒ Выводы
- ⇒ Полезные ссылки

*В этой главе речь пойдет об Интернет-статистике: какие существуют способы получения и обработки статистических данных о вашем сайте, что можно узнать о сайте и его продвижении с помощью статистики.*

### 9.1. Интернет-статистика: цели и задачи

Сайт и реклама в Сети являются, как правило, не единственным средством продвижения и продаж продукта, одновременно используются и другие средства маркетинга — реклама в газетах и телевидении, промоушен-акции и т.п. Точно оценить вклад каждого из способов продвижения трудно, особенно в случае, когда проводится “смешанная кампания” и вклад может быть совместным.

Однако Интернет-технологии позволяют детально изучить поведение привлеченных клиентов: на вашем сайте “все ходы записываются” и могут быть потом проанализированы. Результаты продвижения сайта или рекламной кампании могут оцениваться сразу после ее начала, что дает возможность быстро изменить параметры кампании или скорректировать усилия по поисковой оптимизации.

Для этого и служат средства Интернет-статистики.

Статистика нужна для оценки эффективности сайта и рекламной кампании как инструментов решения конкретных коммерческих задач (обеспечение продаж, технической поддержки продукта и т.п.).

Вот какие задачи можно решать с помощью анализа статистики сайта.

#### **Оценка общей аудитории сайта**

- *Общая статистика*: количество посетителей и просмотренных ими страниц, количество новых посетителей, региональное распределение посетителей, социо-демографические характеристики, сравнение с конкурентами.
- *Источники посетителей для сайта*: откуда они пришли (с поисковых систем, благодаря рекламной кампании), какие поисковые фразы использовались.
- *Детальная оценка посещаемости сайта*: какие разделы сайта популярны, какова средняя глубина просмотра сайта, по каким путям происходит просмотр документов на сайте.

#### **Оценка качества аудитории**

- *Целевая аудитория*: сколько посетителей оказались целевыми, т.е. совершили действия, желательные для владельца сайта (посетили целевые страницы, положили товар в корзину и т.п.).
- *Откуда приходят целевые посетители*, т.е. какая реклама наиболее эффективна для конечной цели сайта.

#### **Технические параметры**

- *Объем, сроки и качество индексации сайта поисковыми роботами.*
- *Наличие технических проблем*: ссылки на несуществующие страницы, перегрузка сайта.

В настоящее время не существует единого средства статистики, которое могло бы решить все перечисленные задачи. В зависимости от ваших потребностей нужно использовать одно или несколько из описанных ниже решений.

## 9.2. Какие параметры посещаемости сайта нужно оценивать

Системы анализа посещаемости сайта предлагают пользователю множество отчетов о посещаемости, которые можно разбить на несколько групп, рассмотренных ниже.

### Суммарные данные о посещаемости

В эту группу входят параметры, описанные ниже.

- 1. Показы.** Количество страниц, просмотренных всеми пользователями. Этот параметр показывает, сколько информационных страниц (текстовых, HTML, ASP, PHP) просмотрели пользователи. В разных системах данный параметр называется по-разному: показ, просмотр, хит, загрузка. В ряде случаев разделяют “загрузки” (просмотры) и “перезагрузки” страницы (повторные просмотры посредством щелчка на кнопке Refresh/Reload в браузере).
- 2. Количество сессий.** Под сессией понимается серия просмотров страниц сайта одним посетителем. Сессия считается оконченной, если в течение какого-то времени посетитель не обращался к сайту.
- 3. Количество пользователей, или аудитория сайта.** Этот параметр показывает количество посетителей, просмотревших сайт за определенное время. Параметр определяется всеми по-разному, почему это так и какие бывают сложности определения размера аудитории, подробнее описано ниже.
- 4. Количество новых пользователей.** Подсчитывается аудитория, посетившая сайт впервые. Этот параметр еще менее точен, чем общая аудитория, в силу сильно разнящихся определений того, что такое новый пользователь. В большинстве случаев данные о новой аудитории получаются несколько завышенными — ниже, в разделе о способах идентификации посетителей, мы подробнее объясним, почему это происходит.
- 5. Количество уникальных IP-адресов (или хостов).** Показывает, с какого количества уникальных IP-адресов просматривались страницы сайта. Данный параметр ранее заменял подсчет аудитории сайта, так как его технически проще посчитать (у каждого пользователя есть какой-то IP-адрес, в отличие от файлов cookie и прочих средств идентификации). Сейчас в большинстве систем статистики этот параметр поддерживается скорее по историческим причинам.

Заметим, что количество просмотренных страниц и количество сессий можно складывать по часам, дням, неделям и т.п., чтобы получить суммарные показатели за какой-то период времени. А вот все остальные параметры напрямую не суммируются — например, у аудитории сайта есть повторяемость, повторные заходы, поэтому эти данные нельзя суммировать напрямую.

## Характеристики аудитории сайта

Для аудитории сайта можно определить несколько параметров.

- **Географическое распределение аудитории.** Для оценки этого параметра используется тот факт, что IP-адреса, как правило, имеют географическую привязку и для каждого пользователя можно определить географическое положение. (Исключение составляют большие распределенные корпоративные сети, как у Газпрома или ЛУКОЙЛа, у которых один и тот же IP-адрес может означать как Москву, так и Новосибирск.) Точность этих данных достаточно хорошая — практически для 100% пользователей можно определить страну, для 95–98% посетителей из России — регион с точностью до субъекта федерации, а в ряде случаев и точнее. Необходимо понимать, что для пользователей, использующих серверы-посредники (проxy-сервисы), — корпоративные или публичные — часто определяется местоположение сервера-посредника, а не пользователя.
- **Активность аудитории.** Под активностью понимается число страниц, просмотренных одним посетителем в течение сессии либо в течение всех сессий за одни сутки. В зависимости от способа подсчета данный параметр может быть как совершенно бессмысленным (когда среднее число просмотренных страниц получается в результате деления числа просмотренных страниц на число посетителей), так и предоставляющим полезную информацию (например, гистограмма распределения длин сессий, количество посетителей просмотревших больше чем  $N$  страниц).
- **Длительность сессии.** Многие статистические системы подсчитывают показатель “время просмотра сайта пользователем” (длина сессии в минутах). Однако этот показатель сложно назвать осмысленным, поскольку неизвестно, когда пользователь покинул сайт. Действительно, невозможно выяснить время просмотра *последней* страницы сессии, а она с большой вероятностью была самой важной. Например, можно предположить, что пользователь в несколько быстрых навигационных кликов добрался до страницы с содержанием и изучал его затем полчаса — так вот эти последние полчаса невозможно “засечь” (потому что потом пользователь мог просто набрать другой адрес в браузере, уже не обращаясь к вашему сайту, или вообще выключить компьютер).

## Ссылающиеся сайты: источники посетителей

При переходах по ссылкам системам статистики обычно бывает доступна информация о странице, с которой пользователь перешел на ваш сайт. Используя эти данные, статистические сервисы предоставляют следующую информацию о переходах пользователей:

- сайты, с которых был выполнен переход;
- точные URL (сайт + страница на сайте), с которых был переход.

(Ну и, естественно, записываемая информация всегда включает дату и время перехода.)

Эти параметры могут быть выражены в “пользователях” (т.е. несколько переходов одного пользователя по ссылке считается за одного “пользователя”) и в “переходах”.

Детальный анализ переходов позволяет также получить данные, приведенные ниже.

- Распределение сайтов-источников трафика по группам (поисковые системы, каталоги и рейтинги, форумы, остальные сайты).
- Для поисковых систем как источников трафика — по каким поисковым запросам осуществлялись переходы. Для получения данного отчета система статистики должна знать поисковую систему и уметь декодировать параметры ее адресной строки (URL) для получения исходного запроса.

## Популярность страниц и разделов сайта

Информацию о посещаемости можно сгруппировать по страницам и каталогам посещаемого сайта. В результате будут получены такие отчеты.

- *Популярные страницы* (документы). Для каждой отдельной страницы можно посчитать количество просмотров и количество просмотревших ее пользователей.
- *Популярные группы страниц* (подкаталоги). Аналогичные метрики для группы документов в одном подкаталоге.
- *Пути посетителей по сайту*. На основании анализа пользовательских сессий можно получить отчет “пути по сайту”, т.е. списки страниц, просмотренных в течение одной сессии. Этот отчет обычно считается достаточно важным, однако на практике обычно получается, что число используемых пользователями путей очень велико, а сколько-нибудь популярными являются только переходы с головной страницы сайта в один из подразделов.

## Целевая аудитория и фильтрация данных

Целевой аудиторией считаются те посетители, которые пришли на сайт с целью, приятной для владельца сайта (что-либо купить на продающем сайте, узнать о продукте на сайте поддержки продаж, ознакомиться с документацией на сайте техподдержки и т.п.). Так как прочитать мысли посетителя обычно затруднительно, целевую аудиторию отличают по действиям, совершенным на сайте. Такими действиями могут быть:

- онлайн-покупка;
- помещение товара в “корзину” или список пожеланий;
- переход на страницу “контакты” или “как проехать”;
- чтение описания товара или загрузка каталога;
- заполнение формы запроса дополнительной информации или подписки на рассылку;
- повторный заход на сайт;
- просмотр определенного количества страниц сайта.

Какое именно действие является целевым, может определить только владелец сайта. Обычно для выявления целевого действия требуются изменения на сайте — например, расстановка меток, создание специальных “приземляющих страниц” для каждой рекламной кампании, размещение купонов со скидками с возможностью печати на принтере и т.п.

Если мы можем отличить целевого посетителя от случайного, то появляются дополнительные возможности по оценке как каналов привлечения посетителей вообще, так рекламных кампаний в частности. Тут можно действовать двумя способами.

- Разделить всю аудиторию сайта по источникам (реклама, поисковые системы и т.п.) и для каждого источника оценить, сколько пользователей совершили целевое действие.
- Выделить из аудитории всех тех, кто выполнил целевое действие, после чего посмотреть, с каких сайтов (рекламных кампаний, поисковых фраз) пришли данные посетители.

Обе методики будут давать близкие результаты с точностью до повторных посетителей (если посетитель приходил на сайт дважды и только один раз совершил целевое действие, то во втором случае он будет учтен, а в первом — нет).

## **Сравнимость результатов для разных статистических систем**

Разные статистические системы по-разному подсчитывают пользователей. Результаты подсчета должны быть примерно одного порядка, однако даже расхождение посещаемости на десятки процентов в разных системах статистики может быть связано просто с различиями в методах подсчета и определения повторности захода, способа вычисления длительности сессии и т.п.

Вследствие этого все характеристики, которые напрямую или косвенно учитывают аудиторию сайта, оказываются несравнимыми для разных систем статистики.

Загрузки страниц всеми системами подсчитываются практически одинаково (в случае разделения “загрузок” и “перезагрузок” возможно сложить эти два параметра), поэтому все показатели, связанные с количеством просмотренных страниц, включая число переходов из поисковых машин и т.п., должны быть

достаточно близкими (случаи потери данных и пользовательских ошибок подробно рассмотрены ниже).

Географические данные рассчитываются по самым разным базам данных, поэтому могут отличаться в разных системах статистики. При этом определение географии с точностью до страны происходит практически со 100%-ной точностью, а для географических баз данных по городам/регионам возможны большие расхождения.

### Устойчивость и разброс данных

Как и большинство других естественных процессов, посещение сайта пользователями носит случайный или околосо случайный характер. В первом приближении к цифрам посещаемости применимы обычные статистические критерии: можно считать, что естественный разброс посещаемости (сайта в целом, отдельной страницы, переходов с поисковика) пропорционален квадратному корню из основной величины.

Например, если в среднем за сутки на сайт заходит 400 человек, то посещаемость в диапазоне от 380 до 420 посетителей в очередные сутки укладывается в 95%-ные рамки “стандартности”. Другими словами, если вчера у вас было 419 посетителей, а сегодня 385, бить тревогу не нужно, если только этот эффект падения не становится устойчивым.

С ростом величины посещаемости эта относительная величина естественных колебаний падает.

## 9.3. Откуда берутся данные для анализа

Все системы статистики берут свои данные из событий, происходящих при обращении браузера пользователя к Web-серверу, который обслуживает ваш сайт.

События эти записываются в журнал сайта, так называемый *лог*.

### Логи сайта

При обращении пользователя к сайту происходит сразу несколько событий.

1. **Запрос страницы.** Клиент (браузер) передает на WWW-сервер запрос на выдачу страницы, при этом на сервере становятся известны:
  - а) *обязательные параметры:* IP-адрес клиента, запрашиваемый URL, дата и время запроса (по IP-адресу можно определить географическое положение клиента);
  - б) *дополнительные параметры* (необязательные, но обычно передаваемые): используемый пользователем браузер, адрес (URL) предыдущей страницы, с которой произошел переход, имя пользователя и пароль для защищенных страниц и т.д.

**2. Выдача страницы.** Сервер возвращает клиенту запрошенный документ (HTML или картинку).

**3. Запись в лог.** Web-сервер записывает данные о произошедшей транзакции в *журнальный файл* (лог-файл).

Как правило, Web-страницы (в этой главе будем называть их *документами*) являются составными, т.е. они состоят из HTML-текста страницы и некоторого количества вставленных в текст изображений. Все они передаются в браузер пользователя по отдельности — на самом деле при запросе страницы происходит несколько транзакций, записываемых в лог по отдельности. Необходимо упомянуть также особенности составных документов.

- Для каждого вложенного документа (изображения, текста во фрейме, стилового файла и т.п.) пользовательский браузер делает *отдельный запрос*.
- Элементы составного документа могут извлекаться не с одного сайта, а с нескольких разных сайтов (Web-серверов). В этом случае адреса (URL) вложенных элементов описываются в основном, внешнем документе. В данном случае запись о запросе страницы появится не только в логах вашего сайта, но и в журнальных файлах этих третьих серверов.

Вот пример нескольких записей журнального файла:

```
62.205.178.34 [17/Jun/2005:00:06:52 +0400]
"/apache-talk/msg05126.html HTTP/1.1" 12854
"http://www.yandex.ru/yandsearch?rpt=rad&text=RFC1945"
"Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.1; SV1; Maxthon; .NET
CLR 1.1.4322)"
"ruid=0000000142B1DAE9000092EA03030F02"

62.205.178.34 [17/Jun/2005:00:06:52 +0400]
"image/color_logo.gif HTTP/1.1" 2649
"http://www.lexa.ru/apache-talk/msg05126.html"
"Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.1; SV1; Maxthon; .NET
CLR 1.1.4322)"
"ruid=0000000142B1DAE9000092EA03030F02"
```

Здесь первая запись фиксирует получение HTML-страницы, а вторая — вложенной в нее картинки. Следует уточнить, что в этом примере оставлены только существенные для изложения поля данных (в порядке следования: IP-адрес, дата, URL, размер документа, URL ссылающейся страницы, название пользовательского ПО и идентификатор пользователя).

Журнальные файлы представляют собой сырые данные, которые необходимо обработать. Качество обработки этих сырых данных (глубина анализа) и определяет качество той системы статистики, которую вы будете использовать.



Заметим, что в логах сайта содержится вся необходимая информация, и никаких других данных о посещаемости сайта в большинстве случаев не существует.

## Подсчет посетителей

По IP-адресам невозможно точно подсчитать количество посетителей сайта. Это связано с тем, что многие пользователи пользуются корпоративным доступом в Интернет и при выходе в сеть получают один и тот же IP-адрес (адрес корпоративного “прокси”, т.е. сервера-посредника). То же самое верно для пользователей некоторых публичных систем доступа в Интернет (домашних сетей, систем телефонного дозвона) — пользователям каждый раз назначаются разные IP-адреса или один на всех.

Для более точного подсчета посетителей и для распознавания того же посетителя в его следующих заходах на ваш сайт в Интернете используется технология cookie (в переводе с англ. печенье; читается как куки), которая заключается в следующем.

- При первом посещении сайта браузеру пользователя присваивается уникальный cookie-идентификатор — по сути, простая текстовая строка с данными Web-сервера.
- При последующих просмотрах страниц сайта браузер, наоборот, сообщает этот идентификатор выдавшему его сайту (*и только ему*). Таким образом, можно отследить как нескольких пользователей, приходящих с одного IP-адреса, так и одного пользователя, приходящего с разных IP-адресов.

На сегодняшний день для большинства сайтов этот метод является единственным способом отследить поведение посетителей сайта (без их регистрации и последующего ввода имени и пароля).

В то же время, согласно современным оценкам, от 10 до 30% пользователей в течение месяца очищают (или теряют) записи cookies. Это и приводит к тому, что оценка количества *новых* пользователей оказывается завышенной, как уже упоминалось выше.

## Точность подсчетов

Реальная жизнь несколько сложнее описанной выше простой схемы. Аккуратный подсчет статистики сайта затрудняется по нескольким причинам.

- Часть пользователей выходит в Интернет с разделением одного IP-адреса между многими пользователями. Это происходит при разделении одного ADSL-соединения между несколькими клиентами, при использовании корпоративных “серверов-посредников” (прокси-серверов) и т.п.
- Часть пользователей меняет свой IP-адрес в процессе работы (например, после обрыва dialup-соединения восстановление может происходить с заменой IP-адреса).

- Используются персональные и корпоративные брандмауэры или “файрволлы” (firewall) и анонимизаторы, которые вообще не позволяют передавать браузеру пользователя записи cookies и другие необязательные данные запроса.
- Часть обращений пользователей к Web-страницам не фиксируется в логах, так как страница на самом деле берется браузером не с самого сайта, а с локального диска пользователя (если она была скачана ранее) или из Web-акселератора провайдера (для ускорения доступа к сайтам многие провайдеры хранят часто запрашиваемые страницы в специальном буфере — кэше).

Вследствие этих причин полученные статистическими сервисами данные отличаются от реальных. Не существует каких-то корректных способов оценить величину отклонения “реальности” от измеренных характеристик, однако принято считать, что эти отклонения в среднем не превышают 5–10%.

В то же время, если аудитория сайта достаточно велика и нет оснований полагать, что она резко изменится (например, после активной рекламы на ресурсах с большой посещаемостью), то можно считать ошибку измерения постоянной и успешно сравнивать между собой данные по соседним дням, неделям, месяцам.

## 9.4. Обработка данных: собственный анализ данных или счетчик?

Существуют два основных способа обработки исходных данных о посещаемости.

- Накопление журнальных файлов у себя и их самостоятельный анализ с помощью *программы анализа логов*. Для такого анализа вам потребуется купить и установить соответствующую программу и прилагать регулярные усилия для получения и обработки данных.
- Использование сервиса обработки данных (счетчика). В этом случае третья сторона — счетчик — ведет журнал загрузок страницы. Делается это с помощью размещения на вашем сайте элемента сайта счетчика (обычно картинки). Затем эти данные анализируются и предоставляются потребителю в обработанном виде.

Каждый из этих способов имеет свои особенности, достоинства и недостатки, сравнительный анализ которых приведен в табл. 9.1. Основные возможности обоих упомянутых выше способов обработки данных приведены в табл. 9.2.

**Таблица 9.1. Сравнение характеристик анализаторов логов и счетчиков**

Параметр	Счетчики	Анализаторы логов
Стоимость установки и эксплуатации	<i>Бесплатно</i> или <i>месячная оплата</i> (5–50 долл. в месяц и выше)	<i>Разовая оплата</i> (иногда несколько сотен долларов) или <i>бесплатно</i> . Встречается также схема аренды ПО (ежегодная оплата)
Трудоемкость установки и эксплуатации	<i>Небольшая</i> : требуется установка кода (фрагмента HTML-текста) на всех страницах сайта (или в общем шаблоне страницы)	<i>Большая</i> : требуется установка программы на сервере, где расположен сайт, либо регулярное копирование лог-файлов на обрабатывающую машину. Для подсчета посетителей может потребоваться дополнительная настройка Web-сервера
Доступность результатов анализа	Реальное время для данных за сегодня, различная по величине задержка для длинных периодов. Подробные данные хранятся не слишком долго	Обсчет требует времени, но для сайтов с посещаемостью в тысячи посетителей в день и менее это время невелико
Доступность исходных данных (лог-файлов)	<i>Низкая</i> : данные накапливаются у владельца сервиса, переход на другой сервис невозможен. Как правило, исходные данные долго не хранятся	<i>Высокая</i> : данные остаются у владельца сайта
Потери данных	<i>Часто</i> : это рядовое явление — при перегрузке сервиса счетчика часть посещений не засчитывается или теряется	<i>Редко</i> : журнальные файлы могут быть утеряны из-за аварии или некомпетентности системного администратора

**Таблица 9.2. Доступные виды статистических показателей**

Параметр	Счетчики	Анализаторы логов
Общая посещаемость сайта	+	+
Посещаемость отдельных страниц и групп страниц	+	+
Переходы по ссылкам с других сайтов	+	+
Трафик с поисковых систем	+/- (детальный анализ поисковых фраз — не для всех счетчиков)	+
Заходы поисковых роботов	- Классические счетчики не отслеживают. Решения этой проблемы для счетчиков имеют свои недостатки	+

Параметр	Счетчики	Анализаторы логов
Сравнение аудитории с конкурентами	+	- (нет данных конкурента для анализа)
Социо-демографические характеристики аудитории	-/+ (возможно для крупных счетчиков, однако такой услуги нет)	-
География посетителей	+	+/- (требуется источник данных о географии)
Фильтрация данных по критериям (источник трафика, точка продаж)	-/+ (возможности ограничены)	+/- (не все анализаторы, хотя возможна фильтрация исходных данных внешней программой)

Обобщая вышеизложенное, основные различия между собственной обработкой данных и сервисом можно свести к следующему.

- Сервис (счетчик) дешевле по стартовым затратам, начать его использовать можно через несколько минут. Получаемые результаты (набор отчетов и их точность) ограничены тем набором возможностей, которые предоставляет сервис счетчика. Обычно у счетчиков есть бесплатные тарифные планы с ограниченными возможностями и платные — с более мощной статистикой.
- Собственная обработка журнальных файлов требует относительно больших денежных и временных затрат на стадии внедрения, однако обеспечивает большую гибкость в ходе дальнейшей эксплуатации.
- Часть отчетов доступна только для одной из технологий.
  - Сравнение аудиторий своего сайта и сайта конкурентов можно сделать только в публичном счетчике. Такой анализ невозможно сделать средствами анализа только собственных журнальных файлов, потому что у вас на собственном сайте нет доступа к данным конкурентов, а у счетчика он есть (конечно, если ваши конкуренты также разместили у себя код данного счетчика).
  - Напротив, счетчики обычно или вовсе не анализируют заходы индексных роботов поисковых систем, или предлагаемые ими решения имеют серьезные недостатки.
- Установка счетчика (особенно многих счетчиков) от третьих сторон может замедлять доступ пользователя к сайту, что в ряде случаев вызывает раздражение у пользователей и владельца сайта.

Таким образом, выбор средств анализа статистики должен производиться, исходя из поставленных задач.

## 9.5. Счетчики: возможности, достоинства и недостатки

### Российские статистические сервисы

Первым российским счетчиком был сервис Rambler's Top100, появившийся в 1997 году. Однако на сегодняшний день предоставляемая этим сервисом статистика не удовлетворяет даже минимальным требованиям. Top100 позиционируется как рейтинг-классификатор и может рассматриваться только как источник целевого трафика, причем только в том случае, если сайт занимает одну из 25–50 верхних позиций в каком-либо разделе рейтинга.

Возможно также использование Top100 как инструмента для сравнения с конкурентами по суточной посещаемости. Делать это следует в случае, когда у конкурента не установлены другие счетчики (напоминаем, что сравнивать данные между разными счетчиками некорректно).

Начиная с 1999 года в Рунете стали появляться счетчики для получения детальной статистики. Рыночная ниша этих сервисов — предоставление аналитики, поскольку они не могут (хотя и пытались) конкурировать с Rambler's Top100 как генераторы трафика.

На сегодняшний день предоставляемый всеми счетчиками набор основных сервисов очень похож, но почти каждый статистический сервис имеет свои уникальные особенности, которые во многих случаях и определяют выбор между ними.

Зарубежные статистические сервисы тоже существуют, однако их применимость в российских условиях ограничена в силу некоторых причин.

1. Анализ поискового трафика требует как минимум поддержки кодировок русского языка и “знаний” о российских поисковых системах. Ничего этого на западном рынке нет. Аналогично, требуется учет хотя бы основных российских каталогов и рейтингов.
2. Для определения географии посетителей желательно иметь географическую базу данных по России хотя бы с точностью до субъекта федерации. Таких данных у зарубежных сервисов нет.

Таким образом, при выборе статистического сервиса приходится выбирать среди российских (и украинских — BigMir) сервисов.

В табл. 9.3 приведены сравнительные данные о функциональности популярных счетчиков. Как можно заметить, среди бесплатных сервисов очевидным лидером является счетчик Liveinternet (в прошлом RAX): набор сервисов и устойчивость у него сравнимы с платными счетчиками-конкурентами.

Таблица 9.3. Сравнительные характеристики счетчиков Рунета

Характеристика	Top.mail	BigMir	Liveinternet	Spylog	Hotlog	Smartlog	Google Analytics
Стоимость	Бесплатно	Бесплатно	Бесплатные и платные сервисы	40 долл. + оплата трафика	6–35 долл. в месяц	10 долл. в месяц и более в зависимости от трафика	Бесплатно с ограничениями по трафику
Данные по общей посещаемости сайта	+	+	+	+	+	+	+
Популярность страниц/групп страниц	+	+	+	+	+	+	+
Пути по сайту	-	Не все пути	-	+	+	Пути из 2 страниц	+(очень подробный анализ)
Точки входа-выхода	-	+	+	+	+	+	+
Ссылающиеся сайты	+	+	+	+	+	+	+
Поисковые фразы	-	+	+	+	+	+	Только Google
Позиции в поисковых машинах	-	-	-	-	-	+	-
Техническая информация	+	+	+	+	+	+	+
География	Страны	Страны, регионы Украины	Страны, регионы России	Страны, регионы России	Страны, регионы России	Страны, регионы России	Страны, регионы (качество географии по регионам — плохое)

Окончание табл. 9.3

Характеристика	Top.mail	BigMir	Liveinternet	Spylog	Hotlog	Smartlog	Google Analytics
Фильтры	-	-	По источникам трафика. Экспресс-аудит — избранные отчеты по сложным фильтрам	Фильтры имеются, набор отчетов по фильтрам ограничен	-	Фильтры по источникам трафика, отчеты ограничены	В терминах “goal” (ключевая страница/ключевой путь)
Надежность, производительность и потери данных	Средняя. В часы пик посещения теряются	Средняя	Высокая	Средняя (для дорогих тарифных планов — отдельные серверы)	Средняя для дешевых тарифов, отдельный сервер для GOLD-тарифа	Неизвестно (сервис новый)	Неизвестно, должна быть высокая
Экспорт данных	-	-	-	+	-	-	+
Представительность отчетов	Средняя	Средняя	Низкая	Высокая	Средняя	Высокая	Высокая

Очень интересный набор сервисов предоставляет Smartlog, однако “молодость” этой системы заставляет относиться к ней с осторожностью — при росте популярности могут начаться проблемы с нагрузками, потерями данных и т.п. (все то, что остальные системы уже прошли несколько лет назад).

### Типичные ошибки пользователей счетчиков

За последние годы нам многократно приходилось слышать, что счетчики считают все неправильно. Очень часто проведенное разбирательство показывало, что проблемы связаны с одной или несколькими типовыми ошибками пользователей, наиболее частые из которых обсуждаются ниже.

#### Установка нескольких (или многих) счетчиков: правило 95%

Очень часто встречаются сайты, обвешанные счетчиками, как новогодние елки, — их может быть 3, 5 и более. Когда пользователь обращается к странице такого сайта, происходит следующее.

- Браузер начинает загружать графические элементы, делая это последовательно (сверху вниз страницы).
- Для каждого счетчика необходимо получить его адрес из системы адресов DNS, установить соединение, получить графическое изображение с внешнего сайта.
- За то время, пока это происходит, пользователь уже может уйти со страницы по ссылке. В этом случае загрузка части счетчиков может не успеть произойти, и данное обращение не будет засчитано.

Экспериментально установлено, что каждый дополнительный счетчик учитывает примерно на 5% меньше посещений, чем счетчик, установленный в HTML-коде выше него.

Устанавливать стоит только те счетчики, наличие которых *необходимо* сайту, это вежливо по отношению к пользователю (ускоряет загрузку страницы), позволяет получать более корректную статистику посещаемости, да и не повышает индекс цитирования статистическим системам в поисковых машинах.

#### Установка кода счетчика внизу (в подвале) страницы

Чем ниже на странице установлен счетчик, тем позднее (в сравнении с другими графическими элементами) будет производиться обращение к нему и тем больше шансов, что клиент до загрузки счетчика успеет уйти с данной страницы на другие страницы вашего сайта.

Для борьбы с этим продвинутые статистические системы предлагают “раздельный” код — “читающий пиксель” (прозрачное изображение 1×1) для установки на самом верху страницы и “логотип” для установки на видном месте. Если вы заинтересованы в получении от счетчика максимально точных показателей, то надо использовать эту возможность.



### Установка кода счетчика не на все страницы сайта

Это очень распространенная проблема — например, если разные разделы сайта собираются из разных шаблонов. Естественно, обращения к тем страницам, где код счетчика не установлен, засчитываться системой статистики не будут, поэтому может возникнуть впечатление, что посещаемость сайта ниже ожидаемой.

### Разные показатели называются одинаково или похоже

Эта проблема является проблемой *интерпретации* данных, однако возникает она достаточно часто. Допустим, в системе статистики А “хитом” называется любое обращение к странице, в системе статистики В — тоже любое, но повторные обращения через 30 и менее секунд не засчитываются, тогда как в системе статистики В не засчитываются перезагрузки (щелчки на кнопке Reload браузера) и повторные обращения менее чем через 30 секунд. Очевидно, что даже по “числу показанных страниц” в идеальных условиях система А покажет наибольшие данные, а В — наименьшие.

### Проблемы счетчиков

Системы статистики, как и любой другой сервис, не идеальны. При этом возникающие у них проблемы с нагрузкой и видимостью в сети моментально становятся проблемами их пользователей, так как при обращении к сайту с установленными счетчиками происходит одновременное обращение пользователя и к системе статистики.

Наиболее распространенные проблемы описаны ниже.

- *Отсутствие связности сети* между пользователем и сервером счетчика. Это может быть временная недоступность или постоянная (например, администратор сети, в которой находится пользователь, запретил доступ к счетчику из соображений приватности либо экономии трафика), либо же сервис счетчика остановлен для регламентных работ. В этом случае клиентский браузер тратит впустую изрядное количество времени на обращение к системе статистики, обращения не происходит, заход пользователя на сайт не засчитан, а пользователь раздражен, поскольку страница загружалась медленно.
- *Перегрузка системы статистики* возникает в ситуации, когда пользователей больше, чем может обслужить система статистики. Все описанные выше симптомы (например, “торможение” при загрузке страниц сайта с установленным счетчиком) повторяются, но происходит это только в часы наибольшей загрузки (с 11–12 до 17–18 по московскому времени). С проблемами перегрузки в разной степени сталкиваются все системы статистики Рунета, за исключением Smartlog (система существует недавно, и пользователей пока мало).

- На практике 10–20% клиентских обращений счетчиками не учитываются; этот процент выше в дни высокой загрузки (будние дни, какие-то события, представляющие массовый интерес) и ниже в дни низкой загрузки.

С учетом вышесказанного следует с огромной осторожностью относиться к предлагаемому рядом статистических систем средствам учета заходов поисковых роботов. Так как роботы не забирают картинки, а вложенные фреймы забирают с опозданием, то единственный способ учета доступа роботов — это вставить на страницы сайта выполняемый код, который будет выполняться при любом обращении к странице и передавать в систему статистики необходимые данные. Если статистическая система “лежит” или перегружена, может получиться так, что и сайт будет недоступен или плохо доступен для индексации поисковой машиной, так как на стороне сервера при каждом обращении к нему выполняется программа (скрипт), ожидающая ответа статистической системы. А это может привести к выпадению сайта из результатов поиска.

Таким образом, мы рекомендуем использовать для учета заходов индексных роботов только системы обработки журнальных файлов, которые работают независимо от сайта, следовательно, лишены ряда проблем счетчиков.

## 9.6. Обзор возможностей анализаторов журнальных файлов

Программы-анализаторы журнальных файлов (логов) предназначены для составления тех же отчетов, что и сервисы статистики, однако они используют те данные, которые накапливаются непосредственно на сайте. Различия в возможностях счетчиков и анализаторов логов были подробно рассмотрены выше.

Как и в случае счетчиков, анализаторам нужно “знание” о поисковых системах и каталогах Рунета, а для анализа поисковых запросов — еще и знание кодировок русского языка. В зарубежных системах статистики это знание, увы, до сих пор отсутствует.

Основные возможности систем анализа журнальных файлов приведены в табл. 9.4, куда включены отечественные решения и три зарубежные системы, две из которых (Analog и Webalizer) бесплатны и поэтому часто включаются в пакет услуг хостерами сайтов, а третья (Webtrends) считается эталонном по набору возможностей.

Таблица 9.4. Сравнительные возможности систем анализа лог-файлов

Параметр	Analog	Webalizer	Webtrends	Статистика Bitrix	NetPromoter (loganalyzer+ статистика сайта)	Spylog Flexolyzer
Стоимость	Бесплатно	Бесплатно	От 895 долл. (Small Business Ed.) до 9995 долл. (Enterprise)	500 долл. (MySQL)-3000 долл. (Oracle). Работает только с сайтами на Bitrix	Статистика сайта: 400 долл., анализатор входит в состав пакета “продвижение сайта”	20000 р. в год
Нагрузочная способность (кол-во обрабатываемых данных)	Высокая за счет низкой функциональности	Высокая (за счет низкой функциональности)	Высокая	Неизвестно	Невысокая, рассчитана на небольшие сайты	
Анализ общей посещаемости	+ (нет подсчета пользователей)	+ (нет подсчета пользователей)	+ (разные способы подсчета пользователей)	+	+	+
Детальный анализ посещаемости по страницам/каталогам	+	+	+	+	+	+
География	-/+ (по доменам)	-/+ (по доменам)	+ (данные по России неточные)	+/- (по странам)	+	+
Статистика по ссылкам	+	+	+	+	+	+
Статистика по переходам с поисковых систем и поисковым фразам	-	-/+ (нет поддержки русского языка)	-/+ (нет поддержки русского языка)	+	+	+

Окончание табл. 9.4

Параметр	Analog	Webalizer	Webtrends	Статистика Bitrix	NetPromoter (loganalyzer+ статистика сайта)	Spylog Flexolyzer
Статистика по заходам поисковых роботов	Общая (выделяется по User-Agent)	Общая	+ (без поддержки части российских поисковиков)	Детальная	Общая в “Статистике сайта”, детальная в Log Analyzer	Детальная, показываются top-N проиндексированных каждым роботом страниц
Точки входа, точки выхода, пути по сайту	-	+ (только точки входа)	+		+	+
Анализ рекламных кампаний	-	-	Развитые средства, чем дороже, тем лучше. Ориентировка на онлайн-продажи	Возможность создавать свои “события” как цепочки посещенных страниц. Простые возможности по анализу рекламных кампаний	Простые средства (заходы по кампании без детального анализа)	Возможность создавать свои отчеты для определенных типов пользователей (посетивших определенную страницу, пришедших по такой-то кампании)
Экспорт данных	-	-	XLS, DOC, PDF	-	-	HTML, XLS, XML
Время на освоение	Минимальное	Минимальное	Большое	Среднее	Среднее	Среднее
Отчеты для руководства (executive summary)	-	-	+ (отчеты очень высокого качества, но на английском)	-	-	-

## 9.7. Выводы

Из вышесказанного очевидно, что возможности по анализу статистических данных возрастают одновременно со стоимостью решения. Как всегда, оптимальный выбор зависит от круга решаемых задач и бюджета.

В то же время оценка реальной эффективности рекламы без детального анализа посещаемости с учетом целевых посетителей практически невозможна, что делает установку того или иного продвинутого решения практически неизбежной. На сегодняшний день это означает либо ежедневный ручной сбор данных из бесплатного счетчика Liveinternet, либо покупку того или иного решения по анализу лог-файлов за несколько сотен долларов.

В следующей главе мы рассмотрим методы анализа предлагаемой на сайте информации. Это необходимо для определения “целей” — страниц, на которых необходимо сосредоточить усилия по продвижению в поисковых системах.

## 9.8. Полезные ссылки

1. Сайт сервиса Интернет-статистики SpyLog (<http://www.spylog.ru/>) является примером крупного портала, посвященного статистическим сервисам.
2. На сайте <http://www.liveinternet.ru/rating/ru/> вы сможете найти рейтинг и статистику для сайтов Liveinternet.ru.
3. Счетчик Rambler's Top100 (<http://top100.rambler.ru/top100/>) является не столько статистикой, сколько самым популярным рейтингом сайтов Рунета.
4. На сайте программы AWLA (<http://www.weblogalyzer.biz/rus/weblogalyzer.html>) вы сможете получить демо-версию анализатора логов и узнать о программе более подробно.
5. В двух частях статьи Юрия Коберского “Особенности национальной Интернет-статистики”, которые можно найти по ссылкам <http://www.searchengines.ru/articles/004597.html> и <http://www.searchengines.ru/articles/004598.html>, вы ознакомитесь еще с одним мнением о различиях счетчиков и лог-анализаторов.
6. Рекомендуем также серию статей Владимира Ринга о всех счетчиках Рунета: “Новейшая обсервация российских счетчиков: формация SpyLOG” (<http://promo.by/searchengines/spylog-overview.html>), “Новейшая обсервация российских счетчиков: формация Rambler TOP 100” (<http://promo.by/searchengines/rambler-top100-overview.html>), “Новейшая обсервация российских счетчиков: формация Rax” (<http://promo.by/searchengines/rax-overview.html>), “Новейшая обсервация российских счетчиков: формация TopMail” (<http://promo.by/searchengines/topmail-overview.html>), “Новейшая обсервация российских счетчиков: формация Hotlog” (<http://promo.by/searchengines/topmail-overview.html>).